

# Rater Agreement on Interpersonal Psychotherapy Problem Areas

John C. Markowitz, M.D.  
Andrew C. Leon, Ph.D.  
Nina L. Miller, Ph.D.  
Sabrina Cherry, M.D.  
Kathleen F. Clougherty, A.C.S.W.  
Liliana Villalobos, B.S.

*There has been much outcome research on interpersonal psychotherapy (IPT) but little investigation of its components. This study assessed interrater reliability of IPT therapists in identifying interpersonal problem areas and treatment foci from audiotapes of initial treatment sessions. Three IPT research psychotherapists assessed up to 18 audiotapes of dysthymic patients, using the Interpersonal Problem Area Rating Scale. Cohen's kappa was used to examine concordance between raters. Kappas for presence or absence of each of the four IPT problem areas were 0.87 (grief), 0.58 (role dispute), 1.0 (role transition), and 0.48 (interpersonal deficits). Kappa for agreement on a clinical focus was 0.82. IPT therapists agreed closely in rating problem areas and potential treatment foci, providing empirical support for potential therapist consistency in this treatment approach.*

(The Journal of Psychotherapy Practice and Research 2000; 9:131-135)

Psychotherapy research historically focused predominantly on treatment process rather than outcome. Interpersonal psychotherapy (IPT), a time-limited treatment for major depression and other disorders,<sup>1,2</sup> has been an exception to that rule. IPT is based on empirical research findings, and its investigators have always stressed outcome research. As a result we know that IPT works, but relatively little about *how* it works. Research has shown that IPT reduces depressive symptoms, but has barely explored the specific choices and interventions that therapists employ to achieve this clinical improvement. Since IPT works, it makes sense to study its workings.

One basic process question involves formulation of the IPT treatment focus. In the early sessions of the 12- to 16-session acute therapy, the IPT therapist must quickly elicit information about recent interpersonal problems from the patient, condense it into a formulation, and obtain the patient's agreement to this formulation, which then becomes the focus of the ensuing

---

Received November 15, 1999; revised March 15, 2000; accepted March 22, 2000. From the Weill Medical College of Cornell University and the Columbia University College of Physicians and Surgeons, New York, New York. Address correspondence to Dr. Markowitz, Weill Medical College of Cornell University, 525 East 68th Street, Room 1322, New York, NY 10021.

Copyright © 2000 American Psychiatric Association

treatment.<sup>3</sup> Although every patient's story differs, IPT formulations link the targeted mood disorder to one of four interpersonal problem areas: grief (complicated bereavement), role disputes, role transitions, and interpersonal deficits. The IPT manual<sup>1,2</sup> provides definitions and examples of problem areas and focuses on observation rather than inference, minimizing subjective factors in diagnosis. Relative to other psychotherapies—and particularly to psychoanalysis, whose richness has made clinician agreement on formulation notoriously difficult<sup>4,5</sup>—the limited options for IPT should facilitate choosing a focus. The question remains: can clinicians agree in diagnosing interpersonal problem areas?

The four problem areas have been paradigmatic of IPT since its inception, and IPT supervision has anecdotally suggested convergent clinician assessment. Yet therapist concordance in diagnosing these problem areas has never been studied. Absence of proven problem area reliability has not evidently hindered IPT efficacy to date, but evaluating reliability might enhance understanding of treatment process and outcome for IPT<sup>3</sup> and for case formulation generally.<sup>5,6</sup> In attempting not a process study but an interrater reliability study to help define this important process issue, we hypothesized that agreement by trained IPT therapists would be high.

## METHODS

IPT problem areas were assessed by three IPT research psychotherapists (a psychiatrist, a psychologist, and a social worker) from a comparative treatment study of dysthymic disorder. All were women; they had a mean ( $\pm$ SD) of  $7.7 \pm 5.5$  years of clinical experience (range 2–13) and  $5.3 \pm 3.1$  years of IPT experience (range 2–8). All three separately rated the same 16 audiotaped initial IPT sessions, and two raters each reviewed two additional tapes, yielding 52 ratings that could be paired for comparison. Data were missing on interpersonal deficits on one rating, and on role disputes on two ratings.

Initial session tapes elicit an anamnesis of the patient's history but predate therapist and patient agreement on a problem area. Three tapes came from another site, from patients with major depression. The remaining 15 were tapes of dysthymic subjects in a randomized 16-week clinical trial who gave informed written consent for treatment and audiotaping. Thus a rater had often herself done the interview on the latter 15 tapes.

Demographics were available on subjects from 15 of the 18 tapes. They were typical of subjects in dys-

thymic treatment trials: 73% were women, of mean age  $44.6 \pm 10.1$  years, with a range from 33 to 60 years. Eighty-seven percent were white, with 1 (7%) Hispanic and 1 (7%) Asian subject; 20% were married, 60% separated or divorced, and 20% never married. Eighty percent had completed at least a two-year college degree; 87% were employed outside the home, and 1 (7%) was a housewife.

Raters used the Interpersonal Problem Area Rating Scale<sup>7</sup> (IPARS; Appendix A), a checklist of previously untested psychometric properties, to rate the presence or absence of each of the four problem areas. They also chose which problem area appeared primary; that is, which they would select as a treatment focus for the patient (IPARS section B.2). When rating patients they had themselves treated, raters were asked to determine problem areas based on the evidence of the single tape rather than on additional knowledge of the patient. Audiotapes spanned the range of IPT problem areas.

The reliability among the three raters was evaluated separately for the classification of each problem area and for classification of the primary problem area, using the analysis of variance approaches to kappa for multiple raters presented by Fleiss.<sup>8</sup> Because problem areas are not mutually exclusive, and because we were interested in rater agreement on each of them, separate kappas were computed for each problem area. Landis and Koch<sup>9</sup> suggest that a kappa above 0.8 be considered "almost perfect," between 0.6 and 0.8 "substantial," 0.4 to 0.6 "moderate," 0.2 to 0.4 "fair," and below 0.2 "poor."

## RESULTS

The three raters detected grief on 10 tapes, role disputes on 20, role transitions on 47, and interpersonal deficits on 9. Multiple problem areas were recorded by at least one rater on 12 tapes (67%). The primary clinical focus was classified as grief on 4 ratings, role dispute on 9, role transition on 38, and interpersonal deficits on 1. The interrater reliability kappas were as follows: grief, 0.87; role dispute, 0.58; role transition, 1.0; and interpersonal deficits, 0.48. Kappa was 0.82 for agreement on the primary problem area that would constitute a clinical focus for treatment. On four tapes, one rater felt that the actual therapist's comments influenced her choice of problem area; on one tape, more than one rater felt this. Deleting the latter tape from analyses produced only trivial changes (kappa: dispute, 0.53, deficits, 0.47).

## DISCUSSION

This study, the first of its kind for IPT, found substantial overall therapist agreement on the presence of interpersonal problem areas. This makes clinical sense: eliciting the history of a death should raise the issue of grief; interpersonal struggles should raise consideration of role disputes; and so forth. Interpersonal deficits, which had the lowest reliability, is the least well defined of the interpersonal problem areas.

Results support the utility of the IPARS instrument, which IPT therapists may find helpful to case formulation. Results also indicate the likelihood of therapist agreement on a key component of the diagnostic phase of the IPT process: the identification of interpersonal problem areas. Moreover, therapists had “near perfect” agreement on the more important clinical issue of choosing a treatment focus for IPT. This second issue depends on the first: had therapists disagreed on the problem areas themselves, they would have had greater difficulty in agreeing on a focus. We infer that therapists clearly agreed on the interpersonal focus of the putative treatment course, and that disagreements were more likely to arise over ancillary interpersonal problem areas: for example, a role dispute peripheral to the patient’s central issue of grief.

This study does not answer the larger question of whether agreement on treatment focus matters—whether the choice of treatment focus actually influences treatment outcome. Presumably it is important that the therapist choose a problem area that the patient finds convincing. That therapists agreed on problem areas in these cases supports the credibility of the focal IPT problem areas. Patients, too, may easily accept these formulations, which would build the treatment alliance and start therapy off strongly.

Limitations of the study include its small sample sizes of audiotaped sessions and raters. The study also assessed tapes mostly of dysthymic patients. Another diagnosis (e.g., major depression or bulimia) or a treat-

ment group for which additional IPT problem areas have been constructed (e.g., depressed adolescents<sup>10</sup>) might affect agreement. The therapist on the audiotape at least occasionally might have slanted the interview, biasing raters and increasing convergence; this is always an issue in reliability studies based on taped interviews. We asked raters to score this very question. That raters occasionally noted the influence of the interviewer on their judgment does not mean that they acted on that influence. Raters try to rate objectively; an awareness of influence—like awareness of transference—may in fact enhance an objective perspective.

That some raters had actually treated the subjects, and hence had knowledge extraneous to the taped session, should only have decreased rater agreement. The study by design tested reliability in a “naturalistic” setting: in all but three cases, tapes were drawn from the study and one of the raters had just been doing her job: getting a history from the patient. The other raters were testing their reliability against that of the actual therapist. Arguably the results support the expected: that given only four general life-event-based choices to select among, trained psychotherapists can agree on clinical choices well beyond what would be expected by chance. Part of the beauty of IPT is its simplicity, for patients and therapists alike.

Research has demonstrated the reliability of core conflictual relationship theme (CCRT) ratings, with kappas similar to those reported here.<sup>11</sup> Studies like these complement outcome research by empirically supporting the mechanics of the therapeutic process. More such research is needed for IPT.

*The authors thank Gregory Hinrichsen, Ph.D., for his generous provision of taped IPT sessions. This work was supported by Grant MH49635 (Dr. Markowitz) from the National Institute of Mental Health, and by a fund established in The New York Community Trust by DeWitt-Wallace. The findings were presented in part at the Society for Psychotherapy Research meeting, Braga, Portugal, June 1999.*

## REFERENCES

1. Klerman GL, Weissman MM, Rounsaville BJ, et al: Interpersonal Psychotherapy of Depression. New York, Basic Books, 1984
2. Weissman MM, Markowitz JC, Klerman GL: Comprehensive Guide to Interpersonal Psychotherapy. New York, Basic Books, 2000
3. Markowitz JC, Swartz HA: Case formulation in interpersonal psychotherapy of depression, in Handbook of Psychotherapy Case Formulation, edited by Eells TD. New York, Guilford, 1997, pp 192–222
4. Seitz PF: The consensus problem in psychoanalytic research, in Methods of Research and Psychotherapy, edited by Gottschalk L, Auerbach L. New York, Appleton-Century-Crofts, 1966, pp 209–225

5. Eells TD (ed): Handbook of Psychotherapy Case Formulation. New York, Guilford, 1997
6. Perry S, Cooper AM, Michels R: The psychodynamic formulation: its purpose, structure, and clinical application. Am J Psychiatry 1987; 144:543-550
7. Markowitz JC: Interpersonal Psychotherapy for Dysthymic Disorder. Washington, DC, American Psychiatric Press, 1998
8. Fleiss JL: Statistical Methods for Rates and Proportions. New York, Wiley, 1981
9. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-174
10. Mufson L, Moreau D, Weissman MM, et al: Interpersonal Therapy for Depressed Adolescents. New York, Guilford, 1993
11. Luborsky L, Diguier L: The reliability of the CCRT measure: results from eight samples, in Understanding Transference: The Core Conflictual Relationship Theme Method, 2nd edition, edited by Luborsky L, Crits-Christoph P. Washington, DC, American Psychological Association, 1998, pp 97-107

#### APPENDIX A. IPT Problem Area Rating Scale (IPARS)

Rater: \_\_\_\_\_ Date: \_\_\_\_\_

Tape #: \_\_\_\_\_

Mark whether each problem area is present or absent, and check ALL appropriate explanatory items. At the end you will be asked to choose a primary focus for IPT with this subject based on the information available from the tape.

##### A. Interpersonal Problem Areas

1. **Grief** present \_\_\_\_\_ absent \_\_\_\_\_  
uncomplicated \_\_\_\_\_ complicated \_\_\_\_\_

If grief is present, identify:

- a. deceased \_\_\_\_\_
- b. relationship to subject \_\_\_\_\_
- c. date of death \_\_\_\_\_
- d. number of months between death and onset of depression \_\_\_\_\_

2. **Interpersonal Dispute** present \_\_\_\_\_ absent \_\_\_\_\_

If present, identify:

- a. significant other \_\_\_\_\_
- b. does an impasse exist? Yes \_\_\_\_\_ No \_\_\_\_\_
- c. predominant theme of dispute:
  - i. authority/dominance \_\_\_\_\_
  - ii. dependence \_\_\_\_\_
  - iii. sexual issue \_\_\_\_\_
  - iv. child-rearing \_\_\_\_\_
  - v. getting married/separation \_\_\_\_\_
  - vi. transgression \_\_\_\_\_
- d. Which theme checked in c. is primary? \_\_\_\_\_

Approximate duration of dispute in months \_\_\_\_\_

3. **Role Transition** present \_\_\_\_\_ absent \_\_\_\_\_

If present, identify:

- a. diagnosis of dysthymic disorder as role transition \_\_\_\_\_
- b. geographic move \_\_\_\_\_
- c. marriage/cohabitation \_\_\_\_\_
- d. separation/divorce \_\_\_\_\_
- e. graduation/new job \_\_\_\_\_
- f. loss of job/retirement \_\_\_\_\_
- g. health issue \_\_\_\_\_
- h. other (specify): \_\_\_\_\_

If more than one checked, which predominates? \_\_\_\_\_

Number of months between event and onset of depression \_\_\_\_\_

**4. Interpersonal Deficit**    present \_\_\_\_\_    absent \_\_\_\_\_

If present, specify characteristics:

- a. avoidant \_\_\_\_\_
- b. dependent \_\_\_\_\_
- c. masochistic \_\_\_\_\_
- d. borderline \_\_\_\_\_
- e. schizoid \_\_\_\_\_
- f. paranoid \_\_\_\_\_
- g. lacks social skills \_\_\_\_\_
- h. other (specify): \_\_\_\_\_

If more than one checked, which predominates? \_\_\_\_\_

**B. Formulation of Therapeutic Task**

1. Rank interpersonal problem areas marked as “present” in order of their apparent impact on the subject’s mood (1 = most important, 2 = of secondary importance, 3 = less important):

Grief \_\_\_\_\_    Dispute \_\_\_\_\_    Transition \_\_\_\_\_    Deficit \_\_\_\_\_

2. Which problem areas would you use to formulate a treatment contract with the subject? (List up to 2, ranking 1 = most important)

Grief \_\_\_\_\_    Dispute \_\_\_\_\_    Transition \_\_\_\_\_    Deficit \_\_\_\_\_

3. What is the rationale for your answer to question 2?

---

---

---

---

---

4. Did the interviewer on the videotape bias your response by indicating his/her opinion of problem areas?    Yes \_\_\_\_\_    No \_\_\_\_\_

5. Did the videotape provide information adequate to formulate a problem area diagnosis?    Yes \_\_\_\_\_    No \_\_\_\_\_

6. Other comments

\*\*\*\*\*

For scoring only: